Strategies for Assessment in Materials Science and Engineering MOOCs: Short-Answer Grading Best Practices

Jessica Sandland[1*] and Philip Rodenbough[2]

[1] Massachusetts Institute of Technology
[2] New York University Abu Dhabi
* Corresponding Author: jgsandla@mit.edu

Abstract

Developing meaningful educational assessments in massive open online courses can be a significant challenge for course developers. In this report, we discuss a variety of strategies that we have used in Materials Science and Engineering MOOC development in order to assure that we are giving our students meaningful and rigorous assessment activities. In particular, this report analyzes and compares short answer quizzes that are self-graded and those that are peer-graded. The analysis highlights the important of careful construction and continuous revision of grading rubrics.

1. Introduction

Massive open online courses (MOOCs) aim to provide meaningful educational experiences to students outside of the boundaries of traditional higher educational spaces. One challenge in creating these online courses is the difficulty of providing rigorous and comprehensive assessment activities to our online students that do not require the direct evaluation by a professor or other subject matter expert - such direct evaluation is typically impossible due to the sheer scale of the student-to-instructor ratio.

This report draws on the work by the Department of Materials Science and Engineering at MIT to develop a comprehensive set of online undergraduate courses that span the depth and breadth of our undergraduate engineering curriculum [1]. To date, our department has produced seven semester-length MOOCs on edX that largely reflect the content offered to our residential university students, and we have two additional semester-length courses in development. These courses can all be characterized as xMOOCs, with instructor-lead, content-based instruction [2].

In order to ensure that we present rigorous problems to our students, we use a variety of tools to create problems that can evaluate the wide variety of skills we aim to teach our students. In addition to the classic multiple-choice question, we also use number or variable input, fill-in-the-blank derivations, drawings, and short open response (or short answer questions) which are either peer-graded or self-graded.

1.1: Numerical or Variable Input

Though Yuan and Powell report that most MOOCs base their assessment on short multiple choice questions [2], we find that in our Materials Science and Engineering courses we often utilize numerical- and symbolic-type problems, which allow our students to work out the solutions to problems and input their answer as either a number or a variable-containing formula. Examples of these two common types of MOOC problems are given in Figure 1.

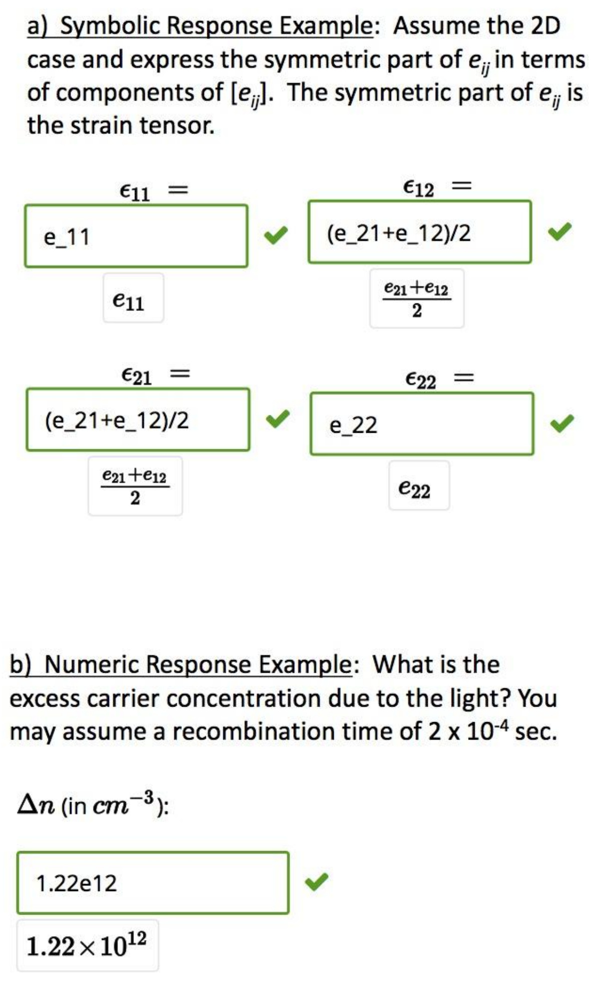


Figure 1a: A symbolic-type problem that allows learners to enter their answer in the form of a formula. This example is taken from 3.072x: Symmetry, Structure, and Tensor Properties of Materials [3]. Figure 1b: A Numeric-Answer problem taken from 3.15x: Electrical, Optical, and Magnetic Properties of Materials [4].

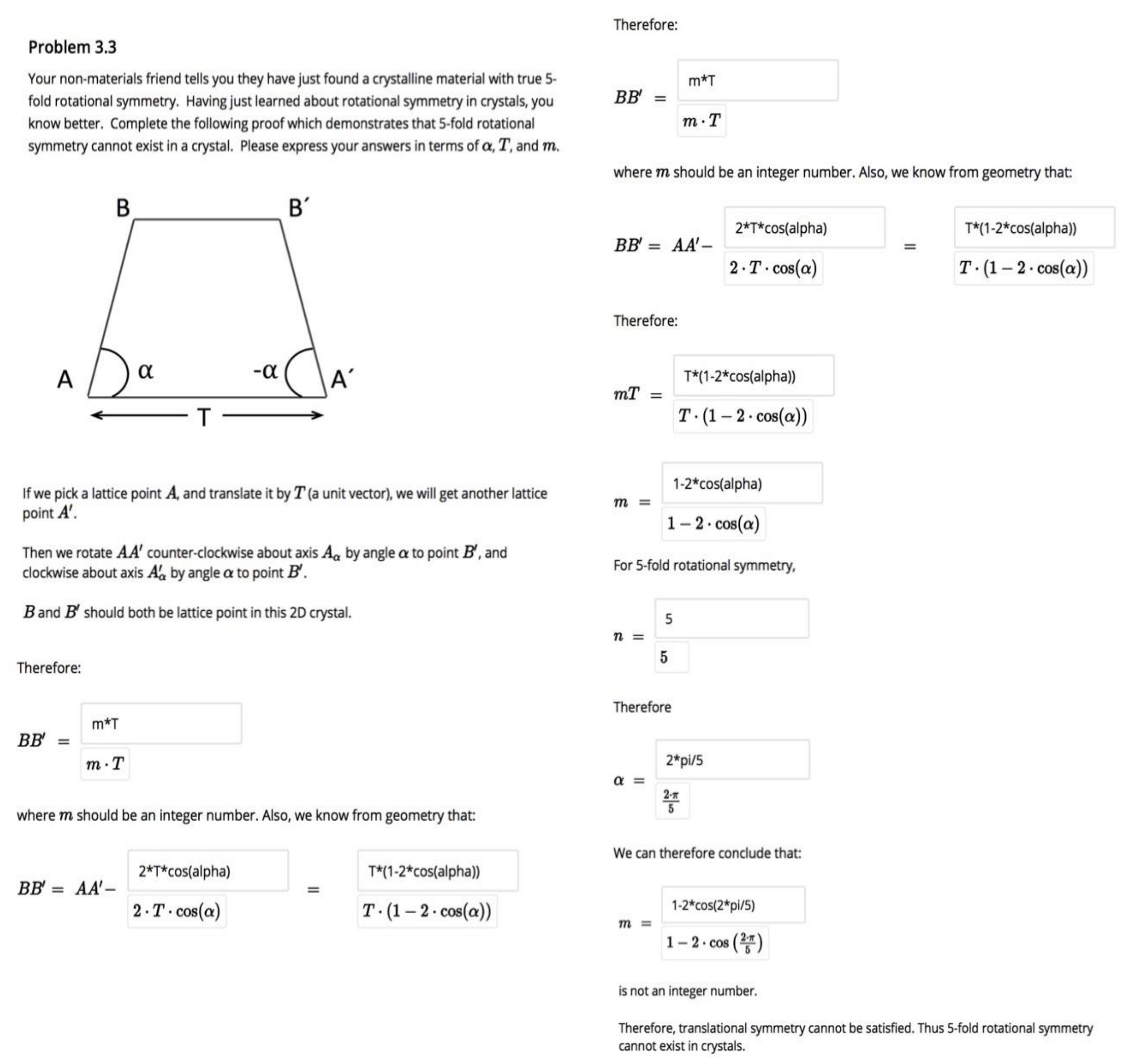
Though we find it valuable to assess a variety of student learning through symbolic and numeric response questions (as well as with some limited use of multiple choice questions), we frequently find ourselves looking for ways to provide more open-ended methods for student response and assessment.

1.2: Fill-in-the-blank derivations

Another important assessment capability for science and engineering instructors is the ability to evaluate student derivations. We have two options for automatically grading this type of problem. First, we can elect to grade only the result of the students' work, having them input a final formula as an answer and checking only that answer. However, this approach can sometimes be problematic for two reasons: first, because if the student makes an error in their final answer we are unable to judge where and how they went wrong in their thinking, and second, because sometimes we ask students to derive equations that are easily looked up in a book, and the process that they use to arrive at their answer is much more important than the correctness of the answer itself.

Our other alternative is to provide students with a skeleton of a derivation, and ask them to fill in the missing information. An example can be found in Figure 2, where students are asked to type the missing information into the text boxes.

In this way, we can guide students through a derivation, checking their understanding at each point. This problem solving structure has similarities to the faded example approach to teaching problem solving - an approach in which students are expected to complete more and more of a problem on their own. This approach bridges the gap between studying worked solutions and being able to solve a problem independently, and has been demonstrated as an effective way of learning during the initial stages of cognitive skill acquisition [6].



Problem 3.3

Your non-materials friend tells you they have just found a crystalline material with true 5-fold rotational symmetry. Having just learned about rotational symmetry in crystals, you know better. Complete the following proof which demonstrates that 5-fold rotational symmetry cannot exist in a crystal. Please express your answers in terms of α, T, and m.

If we pick a lattice point A, and translate it by T (a unit vector), we will get another lattice point A'.

Then we rotate AA' counter-clockwise about axis A_α by angle α to point B', and clockwise about axis A'_α by angle α to point B'.

B and B' should both be lattice point in this 2D crystal.

Therefore:

$BB' =$ m*T $m \cdot T$

where m should be an integer number. Also, we know from geometry that:

$BB' = AA' -$ 2*T*cos(alpha) $2 \cdot T \cdot \cos(\alpha)$ $=$ T*(1-2*cos(alpha)) $T \cdot (1 - 2 \cdot \cos(\alpha))$

Therefore:

$BB' =$ m*T $m \cdot T$

where m should be an integer number. Also, we know from geometry that:

$BB' = AA' -$ 2*T*cos(alpha) $2 \cdot T \cdot \cos(\alpha)$ $=$ T*(1-2*cos(alpha)) $T \cdot (1 - 2 \cdot \cos(\alpha))$

Therefore:

$mT =$ T*(1-2*cos(alpha)) $T \cdot (1 - 2 \cdot \cos(\alpha))$

$m =$ 1-2*cos(alpha) $1 - 2 \cdot \cos(\alpha)$

For 5-fold rotational symmetry,

$n =$ 5 5

Therefore

$\alpha =$ 2*pi/5 $\frac{2 \cdot \pi}{5}$

We can therefore conclude that:

$m =$ 1-2*cos(2*pi/5) $1 - 2 \cdot \cos\left(\frac{2 \cdot \pi}{5}\right)$

is not an integer number.

Therefore, translational symmetry cannot be satisfied. Thus 5-fold rotational symmetry cannot exist in crystals.

Figure 2: An example of a fill-in-the-blank derivation, taken from 3.012x: Fundamentals of Materials Science [5] Students enter mathematical expressions into the boxes to complete the proof.

1.3: Drag-and-Drop Drawings

Another challenge in designing MOOC assessment arises from the difficulty in asking students to provide sketches, drawings and diagrams. It is essential that engineering students be able to create a wide variety of drawings and diagrams to increase and demonstrate their understanding of course material, yet evaluating this work in a computerized fashion can be challenging.

To address this challenge, we often use a drag-and-drop tool to enable students to make drawings and diagrams by essentially constructing them out of their constituent pieces. Students choose the correct diagram components, and drop them onto the appropriate place in the diagram, building up their drawing piece by piece. An example drag-and-drop drawing is shown in Figure 3.

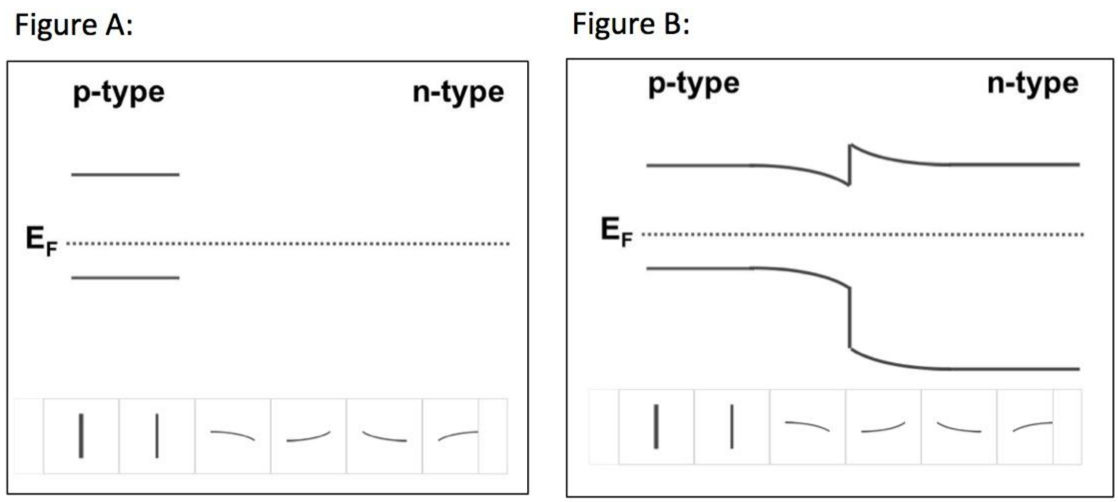


Figure 3: A drag-and-drop style problem from 3.15x: Electrical, Optical, and Magnetic Materials and Devices [7]. Figure A shows the problem as it is presented to the students: a basic outline of the drawing is given above, and a menu of draggable elements are given below. Figure B shows the completed diagram, once all of the elements have been dragged to their correct locations.

Another, related skill we want to develop in our students is the ability to correctly identify and label symmetry elements in various structures. The drag-and-drop drawing functionality can also be applied to these kinds of problems. An example of drag-and-drop used for labeling is shown in Figure 4.

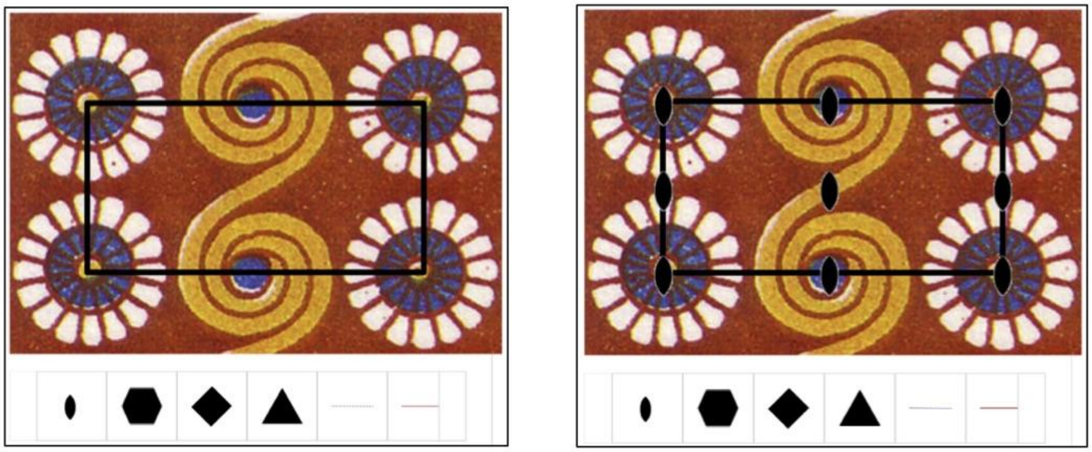


Figure 4: A drag-and-drop style problem from 3.072x: Symmetry, Structure, and Tensor Properties of Materials [3]. The figure on the left shows the problem as presented to students, with a template image above, and draggable elements below. The figure on the right shows a completed diagram, with the correct elements dragged to their proper place. The background image is from Wikipedia and is in the public domain [8].

Other groups are also investigating technologically advanced ways of incorporating drawings into assessment (MITx's Sketch input tool, for example [9]), which provide further opportunity to introduce drawing and sketching into the edX environment.

1.4: Short Answer-Style Quizzes: Self-Graded

One type of exercise that can be particularly difficult to assess in a MOOC environment is the short answer question. In traditionally assessed classes, we frequently ask students to write short explanations, descriptions, or brief analyses on a wide variety of topics. In a computer-graded environment, we can convert these questions into multiple choice-style assessments, but this takes away the need for students to generate their answers in an independent fashion. There are groups working to address some of these issues (such as ETS's C-rater, for example [10]), but many such solutions are currently nonstandard on the edX Studio LMS.

One standard method for grading short-answer style questions is student self-grading. In our Materials Science & Engineering MOOCs, we use self-assessment in two different ways: ungraded self-assessment and graded self-assessment.

Typically, we include ungraded short answer self-assessment questions between lecture video segments as a learning check. Students are asked a question, and are provided with a text box where they can type their ideas. Once students have completed their responses, they click on a “Show Answer” button, and an instructor-created answer is shown to them, allowing the learners to compare their answer to an expert answer. These questions do not contribute to learner grades in any way; their only purpose is to deepen student learning. An example is shown in Figure 5.

Alternatively, we provide learners with similar questions, but after they view the instructor-created answers, they are expected to mark their answer as correct or incorrect, and their evaluation contributes to their final grade in the class.

In-Lecture Exercise 2.2

0 points possible (ungraded)

Silicon carbide (SiC) is a semiconductor used in a few specialized devices. Here are some data in comparison with Si:

	SiC	Si	
Band gap	2.2	1.1	eV
Electron mobility	800	1500	cm^2/Vs
Breakdown field	3.5	0.6	MV/cm
Saturation drift velocity	2×10^7	1×10^7	cm/s

Effective density of states: Similar for SiC and Si

Why do you think SiC might be good for high-voltage and high-temperature devices?

Figure 5: An example of a ungraded self-evaluated in-lecture short answer question from 3.15x: Electrical, Optical, and Magnetic Materials and Devices [4]

1.5: Short Answer-Style Quizzes: Peer-Graded

Instead of utilizing student self-grading, we sometimes use peer assessment to evaluate learner responses to short open-ended questions. However, from a learner perspective, it is difficult to have these short peer-assessed questions interspersed throughout the course, because learners must return to each question after a sufficient time period has passed to evaluate their fellow peers’ assignments. This also introduces a delay in feedback for the learners, making it peer assessment a less desirable option for weekly problem sets in which we prefer to give learners timely feedback regarding their answers.

In order to balance these needs, we often elect to administer short answer-style quizzes once during the run of a course. These quizzes consist of typically 6-10 questions that have very specific answers and that can typically be answered in 1-3 sentences. Student graders are then provided with an instructor generated answer and a rubric that asks them to evaluate if their classmates identified specific points in their responses. An example question and rubric is shown in Figure 6.

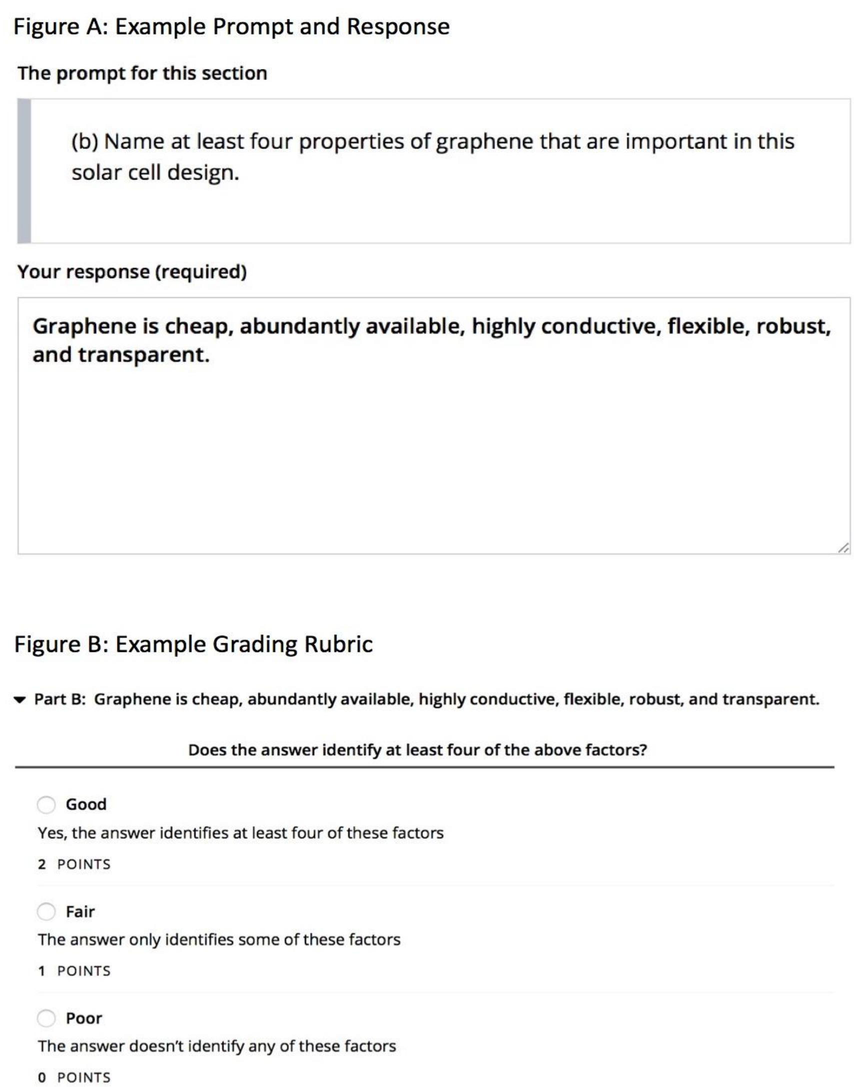


Figure A: Example Prompt and Response

The prompt for this section

(b) Name at least four properties of graphene that are important in this solar cell design.

Your response (required)

Graphene is cheap, abundantly available, highly conductive, flexible, robust, and transparent.

Figure B: Example Grading Rubric

Part B: Graphene is cheap, abundantly available, highly conductive, flexible, robust, and transparent.

Does the answer identify at least four of the above factors?

Good
Yes, the answer identifies at least four of these factors
2 POINTS

Fair
The answer only identifies some of these factors
1 POINTS

Poor
The answer doesn't identify any of these factors
0 POINTS

Figure 6: Example question from a short answer-style quiz in 3.15x. Figure A shows the example question prompt and a potential response. Figure B shows the rubric that learners use to evaluate their colleagues.

1.6: A deeper analysis of self-graded vs peer-graded (vs instructor-graded)

Our primary investigation in this report is an analysis of how successfully students are able to self-review and peer-review fact-based short answer questions. Peer assessment has been demonstrated to have some promise in accurately assigning grades to learners [11]. We will compare the evaluations given by students to staff evaluations to determine how accurately students are able to assess their fellow learners and themselves in this context.

2. Methods

The 2016 offering of MITx DMSE 3.15x on the edX platform was split into three parts: 3.15.1x, 3.15.2x, and 3.15.3x, which covered electronic, optical, and magnetic materials, respectively. A mixture of computer-graded, self-graded, and peer-graded activities were included as assessments across 3.15x. For an in-depth analysis, the authors chose a self-graded activity from 3.15.1x, and a peer-graded activity from 3.15.2x. The self-graded activity asked students to define the terms valence band, conduction band, and band gap. The peer-graded activity asked students to read an article and then reply with short answers to a series of three questions. In both cases, students were provided a rubric after they submitted the answers. Students then graded responses (either their own in the self-graded case, or their peers in the peer-graded case) according to the rubric. For the analysis here, two separate instructors with PhDs and in-depth knowledge of the course material graded all the responses as well.

3. Results & Discussion

The two evaluation methods of short-answer questions (self-grade and peer-grade) showed important differences detailed below.

3.1: Self-Graded Analysis

The instructions given for the activities, along with the model answers provided, are detailed in Box 1. The first three terms were selected for in-depth study: conduction band, valence band, and band gap. Students awarded themselves one point for each definition that matched the instructor's definitions in Box 1.

Please write a brief definition of each of the following terms in the text box below. Then click on the Show Answer button, and evaluate the following: Did your answer match the one in the Solution? (Exact matches are not required, though your answer should contain the main points indicated in the Instructor-generated answer.) After you have evaluated your answer, check the appropriate box below. *Please take care when answering these questions. Though your grade will be based on your own self-evaluation of your answers, your success in the rest of 3.15x depends in part on your understanding of these basic concepts.*
Conduction band: The lowest unfilled energy band in a semiconductor. (Contains no electrons at 0K.) *Valence Band: The highest filled energy band in a semiconductor. (Is completely full at 0K.)* *Band Gap: The energy gap between the top of the valence band and the bottom of the conduction band*
Box 1: Self-graded short-answer questions. Top Box: Instructions given to students. Bottom Box: Correct model answers provided to students after submission of their own responses. Students are then asked to mark their own answer correct or incorrect.

The overall grade frequency for each evaluator type is detailed in Figure 7, and further analysis is provided in Table 1.

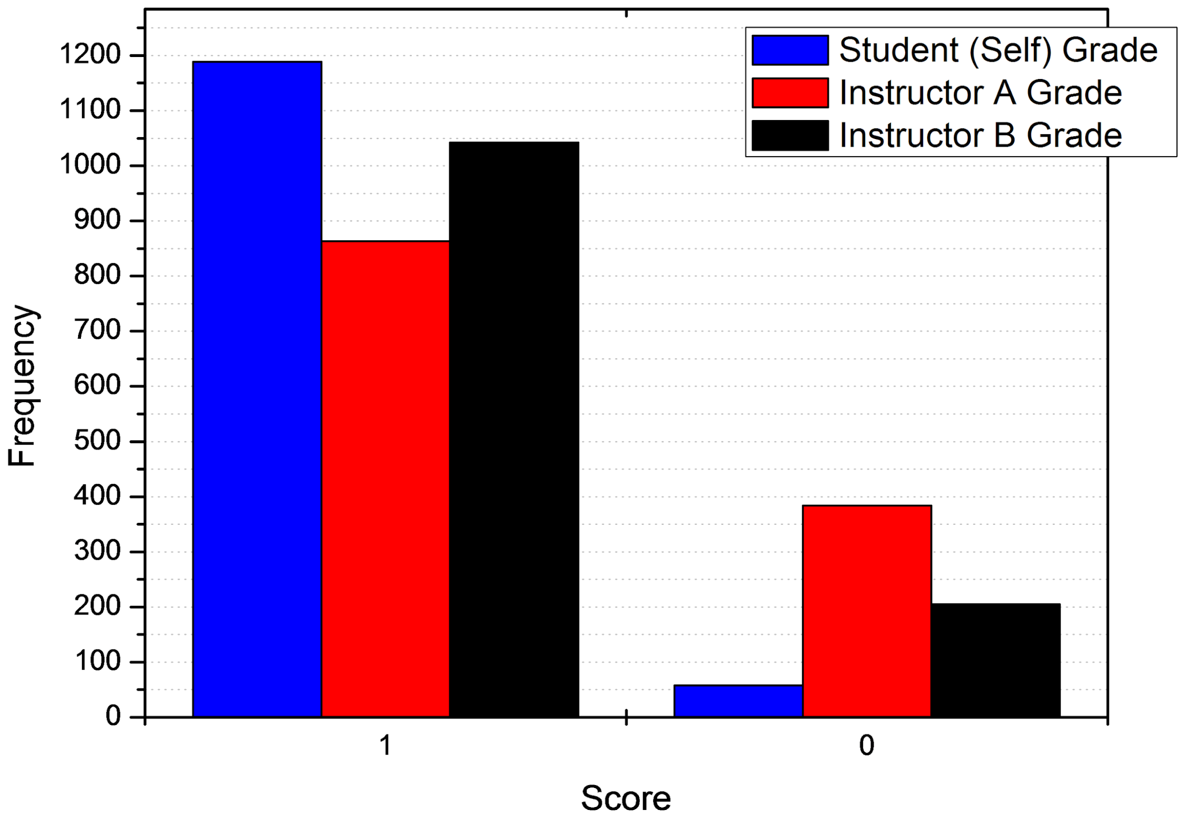


Figure 7: Analysis of a set of responses to questions regarding the definition of conduction band, valence band, and band gap (aggregate total).

Table 1: Additional Analysis of Self-Graded Problems

	Conduction Band	Valence Band	Band Gap
Total non-blank responses	427	421	400
Cases where both instructors gave a grade of 0 but student gave a grade of 1	60 (14.05%)	68 (16.65%)	10 (2.50%)
Cases where both instructors gave a grade of 1 but student gave a grade of 0	4 (0.94%)	1 (0.23%)	4 (1.00%)
Cases where students and both instructors gave a grade of 0	19 (4.45%)	6 (1.43%)	1 (0.25%)

These results are suggestive of several conclusions. One is that students, on average, consistently grade themselves more generously than an instructor would. This is perhaps to be expected.

Another point is that even knowledgeable instructors can differ in their application of a given rubric. For example, many students defined “conduction band” in a manner that was similar to, but less precise than the answer described by the rubric. Should these cases count as correct answers, or as incorrect answers? The best way to evaluate answers such as these is not immediately clear, and the difference in grades between the two instructions suggest divergent judgements. This may be a case of insufficient granularity of a rubric - students are required to give their answers either a grade of 100% or a grade of 0%. In a regular class, most instructor grades would probably just assign (for example) 80% credit to an answer that is 80% correct.

Increasing granularity of a rubric, however, does not always necessarily lead to more consensus - as will be illustrated in the analysis of the peer-graded activities.

3.2: Peer-Graded Analysis

The peer-graded exercises selected for in-depth study came from the final questions for 3.15.2x. The students were asked to read an article entitled *Nanowires and graphene: Keys to low-cost, flexible solar cells* from the Autumn 2013 issue of Energy Futures, the magazine of the MIT Energy Initiative [12]. The questions on the article, along with the correct model answers, are provided in Box 2.

What are the two problems that researchers are trying to solve? *Name at least four properties of graphene that are important in this solar cell design.* *What properties of ZnO nanowires are important for this solar cell design?*
Organic solar cells are less efficient and there is no good choice for electrode material. *Graphene is cheap, abundant, conductive, flexible, robust, and transparent.* *ZnO nanowires increase stability, increase predictability, allow for maximum contact with polymer, and allow electrons to move more quickly to the surface of the device.*
Box 2: A peer-graded exercise. Top Box: Questions asked regarding a provided article. Bottom Box: Model answers shown to students after submission. Students are then asked to grade responses from peers, based on the model answers.

Students were asked to assign a grade of 0, 1, or 2 for each of the three parts in a peer's response. Only the total score (out of 6) was recorded. When the instructors went through and graded responses, the individual part grades were preserved. An analysis of the grading is presented here.

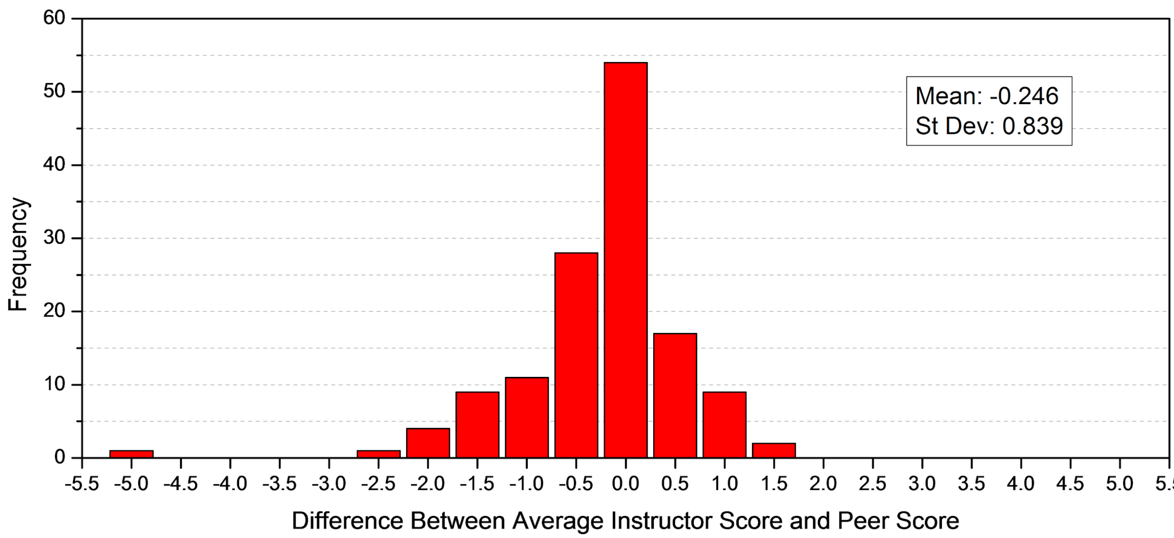


Figure 8: The average instructor score for each response was calculated. Then, that average was subtracted from the peer score. The difference in value is plotted here. On average, out of six points, students rated themselves 0.246 (4.1%) points higher than instructors.

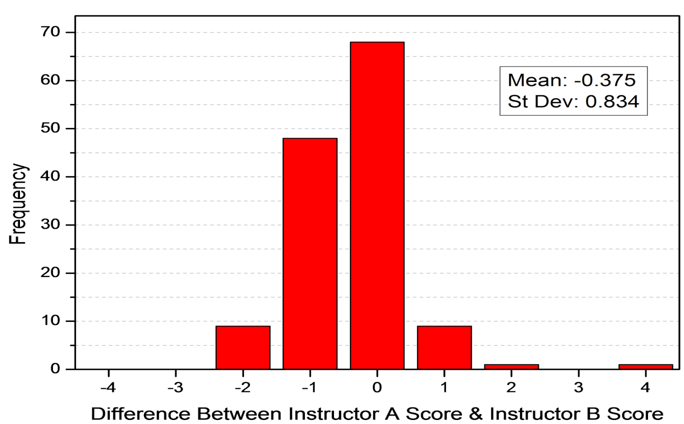


Figure 9: Instructor breakdown of score. The mean difference in instructor scores was larger than the mean difference between instructor average and student-peer scores.

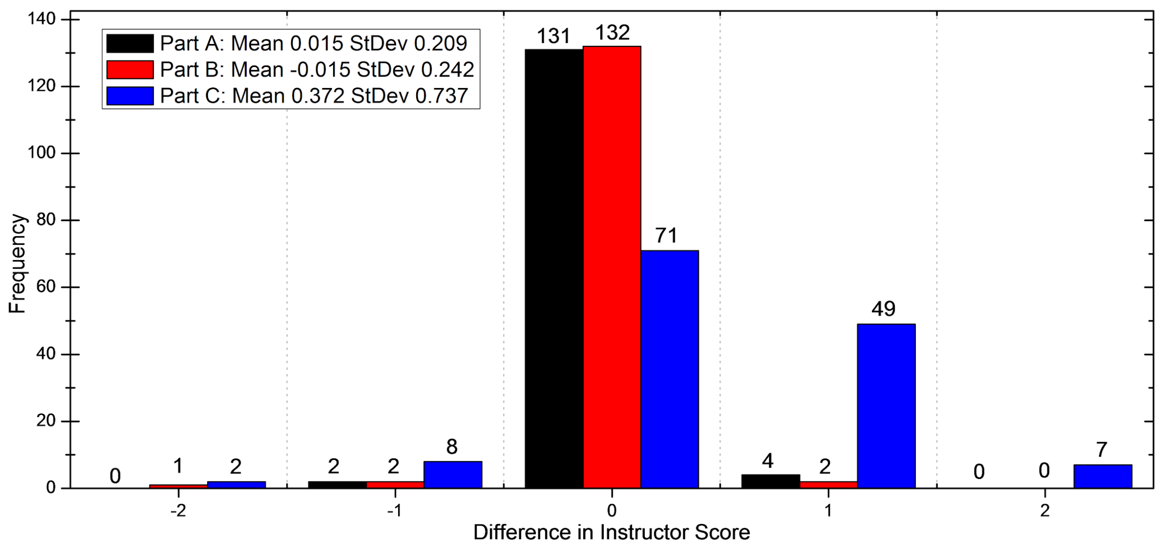


Figure 10: Instructor scores broken down further. Part C was especially controversial.

Table 2: Additional Analysis of Peer-Graded Exercises	
Average instructor score	5.011 (83.5%)
Average peer score	5.257 (87.6%)
Cases where Instr. A, Instr. B, & Peer Agree	50 (36.5%)
Cases where just one instructor & peer agree	51 (37.3%)
Cases where just the instructors agree	19 (13.9%)
Cases where Instr. A, Instr. B, & Peer all graded differently	17 (12.4%)

Table 3: Additional Instructor Grade Breakdown		
	Instr. A	Instr. B
Part A Avg	1.89781	1.883212
Part B Avg	1.89781	1.912409
Part C Avg	1.364964	0.992701

There are several conclusions that can be drawn from this data. First, the overall difference between the instructor grades and the peer grades is relatively small when compared to the difference between instructor grades and student-self grades. This suggest that students grade their peers a little more generously than they should, and that they grade themselves much more generously than they should.

Second, it can be seen that the two instructors interpreted the third question in the rubric a little bit differently. One instructor was more liberal, accepting for partial credit a description of the nanowires themselves rather than the features that made them useful, whereas another instructor was more conservative in following the rubric (which did not allow for such credit). The third question was undoubtedly the most challenging of the three. The first two questions could have allowed students to simply copy and paste from the article, and many did so. The third question required at most an analysis of what features are useful, and at the very minimum, more judicial thought about which passage was to be copy and pasted as an answer. Many students were not so judicial in their response.

4. Conclusion

In our report, we have presented a variety of approaches that we have taken to increasing the rigor and open-ended nature of our MOOC assessments. The data collected and analyzed suggest that both self-graded and peer-graded exercises probably result in inflated (overly-generous) grading – self-graded to a greater extent than peer-graded. The data on the peer-graded responses suggest that a significant portion of the variation in grading may come from how an individual chooses to apply a rubric rather than the quality of the response submitted. This may serve as a worthwhile reminder that it may be important to revise rubrics after reviewing student responses. When students respond in unexpected ways, this may lead to personal judgements in terms of how a rubric may be applied. When rubrics cover the entire range of student responses, with examples pulled from actual student data, this is probably less likely to occur, and hence grading may be more accurate, regardless of the granularity of the rubric, and regardless of whether it is self-graded, peer-graded, or instructor-graded. Such a rubric-revision task, although time-consuming, would not be overly-demanding. Although there are hundreds of student responses, we found that there are in reality only a few kinds of responses that are typically submitted for these kinds of short-answer questions, and the rubrics could certainly be expanded to include nearly the entire scope of student responses. It would be an interesting study to try to confirm these findings with a follow-up analysis of the effects of such revised rubrics.

5. Acknowledgements

The courses described and analyzed in this work represent the efforts of a large number of faculty and online course development experts. In MIT's department of Materials Science and Engineering, courses were developed and created by Prof. Caroline Ross (3.15x), Prof. Eugene Fitzgerald (3.072x) and Prof. Silvija Gradecak (3.012x). In MIT's Office of Digital Learning, Lana Scott, Tsinu Heramo, Douglas McLean, Shelly Upton, Lindsey Weeramuni, Brad Kay.Goodman, David Chotin, Lisa Eichel, and Dana Doyle have provided the course development expertise that has made these courses possible.

6. Bibliography

[1] J. Sandland, "Towards the Development of a Comprehensive Online Materials Science and Engineering Curriculum", *Digital Inclusion: Transforming Education Through Technology: 7th LINC Conference Proceedings,* p. 229-236, 2016.

[2] Yuan, Li, and Stephen Powell. *MOOCs and Open Education: Implications for higher education*. 2013, Centre for Educational Technology & Interoperability Standards, www.researchgate.net/profile/Stephen_Powell5/publication/265297666_MOOCs_and_Open_Education_Implications_for_Higher_Education/links/54081e710cf2c48563b89fbc.pdf. Accessed 22 Oct. 2017.

[3] Fitzgerald, E.A., and Taylor, R. "3.072x: Symmetry, Structure and Tensor Properties of Materials." MOOC offered by MIT on edX. Retrieved October 23, 2017 from https://courses.edx.org/courses/MITx/3.072x/1T2015/course/

[4] Ross, C.A. "3.15.1x: Electronic Materials and Devices." MOOC offered by MIT on edX. Retrieved October 23, 2017 from https://courses.edx.org/courses/course-v1:MITx+3.15.1x+2T2017/course/

[5] Gradecak, Silvija, "3.012x Fundamentals of Materials Science", to be published on edX in 2018.

[6] Renkl, *A, Learning From Worked Examples: How to Prepare Students for Meaningful Problem Solving*. In Applying Science of Learning in Education (2014), pp. 118-130.

[7] Ross, C.A. “3.15.2x: Optical Materials and Devices.” MOOC offered by MIT on edX. Retrieved October 23, 2017 from https://courses.edx.org/courses/course-v1:MITx+3.15.2x+2T2017/course/

[8] Ceiling of an Egyptian Tomb, Retrieved October 23, 2017 from https://en.wikipedia.org/wiki/Wallpaper_group#/media/File:Wallpaper_group-p2-3.jpg

[9] French, J., Segado, M., & Ai, P. Z. (2017). *Sketching Graphs in a Calculus MOOC: Preliminary Results*. In Frontiers in Pen and Touch: Part III: Math for All: High School and College Math. Springer.

[10] Leacock, C. & Chodorow, M. Computers and the Humanities (2003) 37: 389. https://doi.org/10.1023/A:1025779619903.

[11] Piech, C., Do, C, Huang, J., Ng, A., Chen, Z., and Koller, D., “Tuned Models of Peer Assessment in MOOCs”, Proceedings of the 6th International Conference on Educational Data Mining, 2013.

[12] Stauffer, N., “Nanowires and graphene: Keys to low-cost, flexible solar cells”, *Energy Futures*, the MIT Energy Initiative, Autumn 2013.